

---

# **ASSESSMENT OF MAJOR FEDERAL DATA SETS FOR ANALYSES OF HISPANIC AND ASIAN OR PACIFIC ISLANDER SUBGROUPS AND NATIVE AMERICANS**

## **Task 3 Report: Extending the Utility of Federal Databases**

### **Submitted to:**

**Canta Pian  
Human Services Policy  
Assistant Secretary for Policy Evaluation  
Department of Health and Human Service  
200 Independence Avenue, S.W.  
Washington, D.C. 20201**

**OMH-RC-Knowledge Center  
5515 Security Lane, Suite 101  
Rockville, MD 20852  
1-800-444-6472**

### **Submitted by:**

**Westat  
1650 Research Boulevard  
Rockville, Maryland 20850**

**Joseph Waksberg  
Daniel Levine  
David Marker**

---

**May 2000**

## TABLE OF CONTENTS

<u>Section</u>		<u>Page</u>
1	INTRODUCTION .....	1-1
	1.1    Content of Report .....	1-1
	1.2    Subgroups and Databases .....	1-2
	1.3    Task 2 Findings.....	1-3
	1.4    Limitations of Report.....	1-5
2	METHODOLOGICAL AND STATISTICAL ISSUES AFFECTING A SURVEY'S ABILITY TO PRODUCE ADEQUATE PRECISION.....	2-1
	2.1    Definition of "Adequate Precision" .....	2-1
	2.2    Examples of Precision Requirements in Federally Sponsored Surveys .....	2-2
	2.3    Design Effects.....	2-3
	2.4    Consistency in Questionnaire Wording .....	2-6
	2.5    Averaging Over Time .....	2-7
	2.6    Combining Data From Several Surveys .....	2-8
	2.7    Response Rate and Other Quality Issues .....	2-8
	2.8    Estimation of Sampling Errors .....	2-9
3	ABILITY OF CURRENT SURVEYS TO PROVIDE DATA WITH ADEQUATE PRECISION .....	3-1
	3.1    Standards for Precision .....	3-1
	3.2    Effective Sample Sizes Required to Meet Precision Levels .....	3-3
	3.3    Survey Design Effects .....	3-4
	3.4    Nominal and Effective Sample Sizes.....	3-4
	3.5    Surveys and Race/Ethnicity Groups Meeting Standards for Precision .....	3-14
4	COMBINING DATA FOR SEVERAL YEARS.....	4-1
	4.1    Annual vs. Surveys Carried out at Periodic Intervals .....	4-1
	4.2    Maximum Number of Years for Reasonable Analysis .....	4-1
	4.3    Effective Sample Size for Combined Years .....	4-2
	4.4    Surveys Meeting Standards for Precision .....	4-3
	4.5    CPS Labor Force Estimates .....	4-8
5	COMBINING RESULTS FROM SEVERAL SURVEYS.....	5-1
6	SURVEY SUPPLEMENTATION .....	6-1
	6.1    Surveys Requiring Sample Supplementation.....	6-1
	6.2    Designs for Sample Supplementation.....	6-2
	6.3    ACS as a Sampling Frame .....	6-5

## TABLE OF CONTENTS (CONTINUED)

<u>Section</u>		<u>Page</u>
7	SUMMARY OF FINDINGS .....	7-1
List of Tables		
<u>Table</u>		<u>Page</u>
3-1	Effective sample size needed for alternative prevalence rates and levels of precision .....	3-3
3-2	Average design effects for minorities .....	3-5
3-3	Approximations of Hispanic sample cases in the data set .....	3-8
3-4	Approximations of Asian and Pacific Islander sample cases in the data set ...	3-9
3-5	Approximations of American Indian or Alaska Native sample cases in the data set .....	3-10
3-6	Effective sample sizes for Hispanics .....	3-11
3-7	Effective sample sizes for API.....	3-12
3-8	Effective sample size for American Indians or Alaska Natives.....	3-13
3-9	Adequacy of databases for provision of data with acceptable precision .....	3-19
4-1	Effective sample sizes for Hispanic subgroups, using combined years of data .....	4-5
4-2	Effective sample sizes for API subgroups, using combined years of data.....	4-6
4-3	Effective sample sizes for American Indians or Alaska Natives using combined years of data .....	4-7
A-1	Sources used to provide estimates of design effects .....	A-2
List of Appendices		
<u>Appendix</u>		<u>Page</u>
A	SOURCES USED TO PROVIDE ESTIMATES OF DESIGN EFFECTS .....	A-1

## 1. INTRODUCTION

### 1.1 Content of Report

This Task 3 report has two objectives:

1. To examine the ability of the selected statistical databases to provide data on detailed Asian and Hispanic subgroups and on American Indians and Alaska Natives with adequate precision for most practical uses; and
2. To suggest and evaluate methods that could be used to enhance this ability for surveys with insufficient sample size for the provision of reasonably reliable statistics on these minorities.

The report begins with a brief summary of the Task 2 findings; Task 2 inventoried the major characteristics of the surveys and other databases covered in this report which relate to their ability to provide data for the subpopulations. It continues in Section 2 with a discussion of the methodological and statistical issues that need to be taken into account in determining the applicable standards for accuracy, and possible courses of action to achieve this action when current samples are too small to provide estimates with the desired reliability.

Section 3 indicates which surveys meet standards of reliability for at least some of the race/ethnic subgroups. To assist ASPE in determining the conditions for which the data would be useful, we have used several alternate levels of reliability. (In Section 2.1, we suggest some guidelines for choosing what could be considered adequate precision for specific kinds of analyses.) Most of the discussion of adequacy or inadequacy of the data refers to the size of the sampling errors. The sampling errors in a survey depend on both the sample size and the survey's design effects; average design effects are reported in Section 3.3. However, we also point out surveys for which only some of the subgroups are identified on the data file, and consequently, complete analyses of all subgroups are not possible. Idiosyncratic features of some of the surveys that complicate the discussion also are noted.

Sections 4, 5, and 6 describe methods of overcoming the small sample sizes for the detailed race/ethnic groups in most of the surveys. Sections 4 and 5 cover relatively inexpensive methods; combining several years of data for surveys conducted annually in Section 4, and combining results of different surveys for items collected in common, in Section 5. When these two procedures for enhancing the quality are not sufficient, sample supplementation is required.

Section 6 provides information on the sample designs that are efficient for minority supplementation and other statistical procedures that could be used.

Section 7 briefly summarizes the material in this report.

## **1.2 Subgroups and Databases**

For ease of reference, we repeat key items from the Task 2 report, specifically the subgroups of interest, the databases and the appropriate reference dates for the databases. The subgroups of interest are:

Hispanics	Asian or Pacific Islanders	American Indian or Alaska Natives
Mexican-Americans	Chinese	American Indians or Alaska Natives
Puerto Ricans	Filipinos	
Cubans	Japanese	
Central or South Americans	Asian Indian	
Other Hispanics	Korean	
	Vietnamese	
	Hawaiian	
	Other Asian or Pacific Islander	

The databases and the appropriate reference dates are:

	Database	Reference dates
Census	Census 2000 American Community Survey Current Population Survey March Monthly Survey of Income and Program Participation	April 2000 2003, proposed March 1998 Average month 1998 Wave 1, 1996 Panel
NCHS/CDC	National Health Interview Survey National Vital Statistics System Nativity Mortality National Survey of Family Growth National Immunization Survey National Health and Nutrition Examination Survey	1998 1997 1997 1995 1996 1999
AHRQ	Medical Expenditure Panel Survey	1999
HCFA	Medicare Current Beneficiary Survey	Early 1998, 4 Panels
SAMHSA	National Household Survey of Drug Abuse	1997-1998
NCES	National Household Education Survey Early Childhood Longitudinal Survey Birth Cohort Kindergarten Cohort	1996 Year 1, 2000 Year 1, Fall 1998

### 1.3 Task 2 Findings

The Task 2 report contains an inventory of selected major Federal databases, with particular emphasis on information relating to the ability of each database to provide reasonably reliable statistics on the race/ethnic subgroups of interest. The crucial information consists of:

- The population coverage for each survey, the main focus of the content of the questionnaire, and the publication policy.
- Whether the survey or other data collection system currently obtains each respondent's race/ethnic background in the detail required and the question wording used (the question wording and classification detail are expected to change in the next few years to reflect OMB's revised race/ethnic reporting system.)

- The approximate sample sizes for the race/ethnic subgroups of interest.<sup>1</sup>
- Interview policy for the survey.

Appendix B of the Task 2 report contains a detailed description of the way race/ethnicity is currently obtained in each database. The information is summarized below:

- **American Indian or Alaska Native.** All of the databases identify this group. In addition, the census, the U.S. Census Bureau surveys, and many (though not all) of the surveys sponsored by the NCHS distinguish between American Indians and Alaskan natives, but the vital statistics systems do not, nor do those sponsored by NCES. (We note that although the question wording on this racial group has not changed significantly in the last few decades, there are problems of historical comparability. The 1990 Census reported a sharp increase in the number of American Indians or Alaska Natives over the number in the 1980 Census, much greater than can be accounted for by natural increase. Most demographers attribute this change to a heightened interest of persons of American Indian ancestry to acknowledge an affiliation with this racial category.)
- **Hispanic subgroups.** All databases except MCBS and NHES identify each Hispanic person as Mexican-American, Puerto Rican, Cuban, or other Hispanic. In the census, ACS, CPS, and SIPP the “other Hispanics” are further classified as Central American, South American, or other Hispanic. A combined Central and South American classification is used in NVS.
- **Asian and Pacific Islanders.** Considerable variation exists in the way APIs are asked to describe themselves and in the detailed groups that are identified. The decennial census, ACS, NHIS, NHANES, MEPS, NHSDA, and ECLS-K obtain the full level of detail. In the National Vital Statistics Systems (both natality and mortality), all states classify Chinese, Japanese, Hawaiian and Filipinos separately. Vietnamese, Asian Indian, Korean, Samoan, and Guamanian are also separately identified in states that contain about two-thirds of the population in these groups; in the remainder of the U.S. they are combined into an “all other API” category. Since the ECLS-B sample is based on birth registrations, the same classifications are available. CPS, SIPP, NSFG, NIS, and NHES simply identify APIs as a single group, without any further detail. MCBS separates Native Hawaiian or Pacific Islanders from Asians, but does not obtain any further detail. This breakdown used by MCBS is consistent with the recent OMB Guidelines (see Appendix C of the Task 2 Report) that will be adopted by all surveys over the next few years. For simplicity we continue to use the term API in this report.

---

<sup>1</sup> The sample sizes are used to estimate the sampling errors that are applicable to subgroup analysis, and thus to a determination of whether subgroup data for each survey can be obtained with a reasonable degree of precision. The effective sample size, in which the actual sample sizes as shown in Tables 3-3 to 3-5, are deflated by the design effect, is a better guide to the sampling error. Section 2.3 of this report contains a discussion of design effects and Tables 3-6 to 3-8 show average effective sample sizes.

## Limitations of Report

This Task 3 report is essentially limited to the effect of sampling errors on the reliability of the various databases, and possible methods of improving precision when sample sizes are inadequate. There are, of course, other factors that affect the quality of surveys and data files. A complete discussion of these factors is beyond the scope of this report but we wish to call particular attention to several specific issues:

- The reports are intended to be a general reference to a potential audience of analysts and policy makers seeking information on the possible use of these data bases as a source of data on the race/ethnic group of interest, rather than as a technical handbook. We suggest that users, who are not thoroughly familiar with the content of the database being considered and the procedures involved in data collection and data processing, seek appropriate technical assistance from the staff of the relevant agency or from documentation of the survey methods. There are two items that should be particularly examined.
  1. Are the sample sizes shown in tables 3-3 to 3-5 still applicable, or have there been important modifications made in the survey's sample design. We note that small changes in sample size of the order of 10 or 15 percent will have only a negligible effect on the conclusions drawn in this report, and they can be ignored. Important changes in the sample, however, should be taken into account.
  2. What is known about sources of errors in the data, including those arising from possible problems in identifying the race/ethnic groups, respondents' lack of information on some of the subject matter items or misunderstanding of various questions, and potential effects of nonresponse. For example, NCHS studies indicate there may be important issues in death rates for Hispanics, Asian and Pacific Islanders, and American Indians and Alaskan Natives due to misunderstanding of the race question on death certificates or in the censuses and surveys used as the denominators of the death rates. Similar reporting errors and differential nonresponse could affect other statistics.
- It is possible to think of sampling errors in a somewhat broader sense than the term is used in this report. Statisticians distinguish between descriptive and analytic uses of survey data. Descriptive uses provide a profile of a finite population, the population that existed during the period of data collection. Analytic uses occur when survey results examine a process, frequently a "cause and effect" relationship, with the population at the time of data collection considered as a sample of an infinite population. The particular year for the time periods of the study can be considered a single observation from a stochastic process, with neighboring years reflecting additional observations, (for a few years, before long term trends disrupt this model of behavior). NCHS views birth rates as subject to stochastic variation. Similarly, analytic uses would include examination of the effect of

educational attainment on income, the relationship of obesity to various health conditions, etc.

Stochastic processes are subject to sampling errors arising from the erratic variations over time of the statistics studied. In most of the data bases examined for this report, the effect of this source of variation will be trivial compared to the sampling errors due to the sample sizes for data collection. However, the NVS and the Census short forms do not have any sampling errors, but their analyses are subject to a small amount of stochastic variation. NCHS has carried out studies of their effects on birth and death rates, and more detailed information can be obtained from the agency. We note that this Task 3 report is restricted to limitations of the data due to sampling error.

- Information from several sources, each of which is subject to sampling errors and/or other limitations, often is combined for analysis. For example although the numerators of birth and death rates come from vital statistics records that are not subject to sampling errors, the denominators are derived from census reports; some of the census data are based on sample surveys, and others on extrapolation of census data to intercensal time periods. This report does not deal with such special situations, but users who anticipate such analyses should take the more complex sampling into account and, if necessary, seek advice from the agency technical staff.

## 2. METHODOLOGICAL AND STATISTICAL ISSUES AFFECTING A SURVEY'S ABILITY TO PRODUCE ADEQUATE PRECISION

### 2.1        **Definition of "Adequate Precision"**

The level of precision that is considered adequate for a survey should reflect the kinds of analyses to be made and the effect that errors in the statistics are likely to have on practical uses of the information. When differences of 10 or 20 percent will have trivial effects on policy decisions, fairly large sampling errors are tolerable. In other instances, even small errors could have an important adverse effect on uses of the data. The situation is further complicated by the fact that although precision may be satisfactory for most statistics relating to the total population of the subgroup, it may be inadequate for subdomains such as age-sex subgroups, low-income persons, persons in each region of the U.S., etc. It is frequently found that no matter how large the sample for a particular survey, there will be some desirable analyses for which the sample is insufficient. Some examples are separate studies of babies, teenagers, or the elderly; examination of data for persons with income below the poverty level; the rural population; etc.

Consequently, there is no simple or single standard of reliability that is applicable to all studies that may be carried out. Although most large surveys have multiple objectives, in each case, the principal uses of the data should be considered along with the consequences of errors in the data. An important part of the consideration should be whether there is a need for special treatment of certain subdomains. In Section 2.2 we give examples of how some of the major U.S. surveys approach the problem. The budget that is likely to be available should, of course, also be taken into consideration. Another factor that may play a role is the existence of significant nonsampling errors in the data collection system, since there is no reason to incur the cost of a very large sample size if the main quality problem is poor reporting by the respondents rather than sampling error. There is no point in establishing unrealistic standards that cannot be achieved. Section 2.2 contains a few examples of how these and other considerations have been instrumental in establishing standards for some of the major U.S. multipurpose sample surveys, which in turn have determined the sample sizes.

## 2.2

### Examples of Precision Requirements in Federally Sponsored Surveys

Common standards for precision do not exist for U.S. Government surveys. Each survey is viewed as a unique data system designed to meet specific needs. In most cases, the sample size is determined by agreement on the key analytic requirements of the survey weighed against available funding or a realistic budget, rather than to satisfy abstract notions of analytic goals. In a few cases, Government agencies have articulated the principal analytic and policy uses expected of the data and the sampling errors that would permit these analyses. Some examples follow:

- **Sample used in long form of U.S. Censuses.** The 1990 long-form sample contained about 16 percent of the U.S. population and this proportion was repeated in Year 2000. Two factors led to the use of sampling and, in particular, to the choice of 16 percent. First, the existence of considerable nonsampling errors arising from coverage problems and reporting errors by respondents and interviewers indicated that, beyond a certain point, sampling would have only a minor effect on the overall quality of the data. Secondly, a crucial goal of the census is to provide data for small areas, such as tracts, small towns and villages, etc. A 16 percent sample was viewed as mostly satisfying that goal. However, income distributions are frequently used by states and the Federal Government for fund allocation among municipalities, and a 16 percent sample appeared to be inadequate for this purpose for very small municipalities. The sampling rate in these small places was increased, to a maximum of 50 percent.
- **Current Population Survey (CPS.)** Although CPS is a general-purpose social and economic survey with emphasis on labor force characteristics, statistics on unemployment are viewed as particularly important, since movements in the monthly unemployment rate can have a major effect on economic policy. Two aspects of unemployment data determined the monthly sample size for the CPS. First, it was considered necessary to detect a month-to-month change in the national unemployment rate of 0.2 percent as being statistically significant at the 95 percent confidence level. Secondly, it was desirable to have reasonably precise data (level of precision not specifically defined) on unemployment rates for blacks and Hispanics, including separate data for minority youths.
- **National Health and Nutrition Survey (NHANES.)** Since health characteristics are strongly associated with age and sex, the major analyses of the data involve age-sex breakdowns. In addition, as is the case in many other national surveys, separate data are considered necessary for the dominant minority groups; blacks and Mexican-Americans in the case of NHANES. A set of 52 age-sex-race/ethnicity classes was defined and a major goal of the sample design was to provide about the same level of precision for each of these 52 classes. The precision specifications for NHANES III were:
  - A prevalence statistic of 10 percent should have a relative standard error (RSE) less than 30 percent; and
  - Differences of at least 10 percent in health or nutrition statistics between any two classes should be detected with a type I error of no more than 0.05 and a type II error of no more than 0.10.

A sample size of 560 was determined to satisfy these requirements in most classes. Many of the Mexican-American age-sex classes had higher design effects than other classes and needed somewhat larger samples.

- **National Health Interview Survey (NHIS.) Precision Specifications in Considering the Feasibility of Producing Data for States.** The National Center for Health Statistics (NHIS) contracted with Westat to examine the feasibility of producing state data from the NHIS, and to propose methods of enhancing the feasibility for states that currently do not satisfy the precision standards. NCHS specified several alternative precision requirements for this project to give the agency information on what could be accomplished at various budgeting levels. The requirements were:

To achieve specified levels of precision (described below) for four crucial statistics:

- Percent of the total population without health insurance;
- Percent of persons under 19-years of age without health insurance;
- Percent of low-income children under 19-years without health insurance; and
- Percent of persons 18-years and over who are smokers.

- To achieve the same level of precision for five “generic” prevalence rates: 0.01, 0.05, 0.10, 0.15, and 0.20, for statistics that are likely to have design effects of 1.0, 1.5, 3.0, and 6.0.
- Three levels of precision are specified for the statistics described above: coefficients of variation (CV) of 30 percent, 20 percent, and 10 percent.

The generic prevalence rates are a useful model for an examination of the feasibility of producing minority subgroup statistics, and we will focus on these specifications later in this report. However, as stated earlier, no single set of specifications is likely to meet all conceivable analytic needs.

### 2.3 Design Effects

The simple formula for the variance of a sample mean given in most elementary statistics textbooks is  $\sigma_{\bar{x}}^2 = (1 - f) \frac{\sigma^2}{n}$ . For a proportion or a prevalence rate, this formula is equivalent to  $\sigma_p^2 = (1 - f) \frac{p(1 - p)}{n}$ .

In these formulas,  $f$  is the sampling rate,  $n$  is the sample size,  $\sigma^2$  is the population variance of the characteristic being estimated, and  $p$  is the proportion that is estimated. These

formulas apply to the simplest type of situation, that is, use of simple random sampling with all members of the population sampled at the same rate.

In practice, it is rare for population surveys to use simple random sampling. Where interviewing is done on a face-to-face basis (as distinct from telephone or mail data collection) some form of clustering is almost always used to reduce the cost of interview travel. The clustering frequently results from several stages of sample selection, e.g., counties, groups of neighboring households, and members of sample households. Even when telephone or mail is used (e.g., planned as the dominant data collection methods for the census long forms and the ACS); persons within the sample households constitute clusters. The extent to which characteristics of persons within these levels of clustering tend to be correlated influences the size of the sampling variances. In most cases, clustering increases the sampling variances above what would result from a simple random sampling with the same sample size. Variances will also be increased if the sampling rates vary among members of the population. This can come about if some groups are oversampled or undersampled. It can also result from a fairly common practice of selecting a household sample, then choosing one member at random for a more detailed interview. Persons in large households then have smaller probabilities of selection than persons in smaller households. In attempts to compensate for such features of the sample design, statisticians apply devices that tend to reduce variances, principally stratification and sophisticated weighting methods. However, the features that tend to increase variances usually dominate.

The design effect is a measure of the extent to which the interactions of all such features affect the sampling variances. It is defined as the factor by which the variance of an estimate is changed through departure from simple random sampling. It is generally expressed symbolically by  $d$ , so that the variance of a mean becomes  $\sigma_{\bar{x}}^2 = d(1-f)\frac{\sigma^2}{n}$ . As indicated above,  $d$  is mostly, though not necessarily, greater than 1. The value of  $n/d$  is frequently referred to as the effective sample size, since replacing  $n$  by  $n/d$  in the expression for  $\sigma_{\bar{x}}^2$  permits one to use the formulas for simple random sampling.

Design effects differ greatly from one survey to another, since there are important differences among sample designs. They can also vary among different items measured within a survey, and sometimes among specific population groups. A few examples of such variations are described below:

- The CPS provides estimates of the distribution of employees by major industries, as well as measures of the total number of employed persons. For most major industrial groups such as retail trade, manufacturing, etc. there is a rather modest effect of clustering. However, the design effect for agricultural employment is quite large due to the fact that there is a heavy concentration of farming activities in some counties, and practically none in other counties. Furthermore, within counties with extensive agricultural operations, persons living in clusters of households in open country will tend to work in agricultural operations whereas those in towns are more likely to work in support services. The design effect for agricultural employment is therefore likely to be much larger than for most other industry groups. (The CPS employs a special procedure in weighting, referred to as first-stage weighting, to reduce the design effect for rural and farm statistics, but they are still greater than for most other items.)
- The NHIS oversamples blacks and Hispanics in order to improve the reliability of statistics on these two minorities. However, in order to keep the cost of the oversampling within reasonable limits, the oversampling is restricted to the blacks and Hispanics who live in concentrated minority areas. The variation in sampling rates between minorities who live in concentrated areas and those in more integrated areas contributes to the design effects for all statistics on these minorities, as well as increasing the design effects for data on the total population. Clustering, particularly of sample persons within households, also adds to the design effect for some items but not for others. For example, in most households, either all household members have health insurance or none are insured. Obtaining this information for all household members provides only slightly better reliability than if only one person in each household was in the sample. The design effect for this statistic is thus very high. However, estimates on the incidence of health insurance for children or women do not increase the design effect to the same degree because the number of adult women or children in a given household is generally much smaller than the total household size. Similarly, estimates of persons with specific chronic diseases or those with recent hospitalization episodes will have fairly low design effects since there is hardly any clustering. Design effects among items and subpopulations in NHIS range from 1 to 6 so that averages of design effects can only be considered a simplified generalization.
- Both NHANES III and current NHANES oversample Mexican-Americans and blacks in order to permit separate analyses of these two race/ethnic groups. The goals of both NHANES III and current NHANES were to produce reasonably reliable statistics for a set of 52 age-sex-race/ethnicity groups; 14 for both Mexican-Americans and blacks, and 24 for whites and all other persons. This implied starting with a sample designed to provide approximately equal size samples in the 52 groups. Superimposed on this was an oversampling of Mexican-Americans in geographic areas containing high concentrations of Mexican-Americans. The combination of these two features of the sample resulted in sampling rates for Mexican-Americans varying in range from 7.5 to 1. Asian and Pacific Islanders were sampled at the same rate as white persons and their sampling rates varied in a range of 20 to 1. The range among all age-sex-race/ethnicity groups was over 120 to 1. This diversity in sampling rates contributed significantly to the design effects for NHANES and the effective sample size is much lower than the nominal sample size, which was already very small for the API's.

In spite of the diversity in sampling rates, the NHANES sample sizes provide data with fairly good precision for Mexican-Americans in each of the 14 age-sex domains designated

by NCHS for separate analysis. On the other hand, the sample size for even the total API population and for American Indian or Alaska Native is quite low, and it was trivial for individual age-sex groups of these subpopulations.

As the illustrations above indicate, most surveys are subject to a wide array of design effects. If there are a few key statistics in a survey whose importance dominates the analyses and uses of the data (as in the case of unemployment for CPS), then it is useful to concentrate on these statistics in assessing the reliability of the survey estimates. Otherwise, it is sensible to use an average design effect, about midway between the upper and lower levels that are likely to occur. We will follow this latter practice in assessing the ability of the various data sets to meet the needs for data for the different subpopulations of Hispanic and API populations and for American Indian or Alaska Natives.

## **2.4        Consistency in Questionnaire Wording**

Some of the detailed subgroups are not identified on all of the data tapes. Section 1.3 identifies those subgroups that are not fully described on the data records, and indicates whether the omitted subpopulations were not identified in the interview or, if obtained, were not entered into the data tape. In addition, there currently are slight variations among surveys in the way the race/ethnicity questions are worded but, except for the birth registration system, the surveys appear to be reasonably consistent in their classifications. NHANES III probed more intensively than most other surveys to identify persons whose ancestors migrated from Mexico, even though, at this time, they do not consider themselves Mexican-Americans. However, the intensive probing was dropped for current NHANES. It also is important to note that birth certificates ask for race/ethnicity of mother and father, but not of child. Since information for the father is less likely to be available, published data on births are tabulated by the race/ethnicity of the mother, which can introduce some error into the calculation of rates where the numerators are from the NVS-Natality files, but denominators are drawn from other data systems. This inconsistency, however, is not present for infant mortality rates where the linked birth/infant death data set is used and data are tabulated by the mother's race.

The differences among surveys are expected to largely disappear when the revised OMB standards for collecting race/ethnic data, which permit respondents with mixed ancestry to choose more than one race, are implemented. Most surveys will be converting to the new standards over the course of the next 4 or 5 years. The new classifications will bring greater consistency among surveys. NHIS has been collecting data on multiple race identification since

1982, and NHANES currently follows the NHIS approach. To reduce possible problems of historical comparability, NCHS, BLS, and the U.S. Census Bureau have carried out research on strategies to bridge the changes created by the shift to the new classification of race/ethnicity. The proposed OMB revisions in the standards for the federal collection of race/ethnicity are shown in Appendix C, Task 2 Report.

## **2.5 Averaging Over Time**

Section 4 of this report discusses the possibility of improving the precision for some of the subpopulations by combining data for several years. An immediate question is how many years can be combined without seriously affecting the usefulness of the data.

As with so many of the other issues that have been raised, there is no single time period that would be uniformly acceptable for all surveys, or for that matter, for all items within some of the surveys. We suggest that the decision on the number of years to be combined be based on how slowly or quickly the characteristic(s) that are measured in a survey change over time. For example, it is unlikely that there will be dramatic changes over the course of a few years in most of the health or nutrition items covered in NHANES, e.g., prevalence of hypertension, high cholesterol levels, obesity, etc. This, of course, is the reason that NCHS has been comfortable in having previous NHANES data collection extend over a 6-year period. Even though each year of the current NHANES will be based on a random sample, there is no reason why 6 or more years cannot be combined for analyses of data for small population subgroups. Fertility patterns also are likely to change only slowly over time. However, since the NSFG is currently carried out intermittently (about every 5-years), some thought would have to be given to whether combining two cycles of NSFG would excessively stretch the ability to describe the current situation. On the other hand, the limited information on fertility collected annually in CPS could probably be combined over a 3 or 4-year period without any harm, as could the data on educational attainment. Economic statistics, however, can undergo strong fluctuations over a few years, or even less, since they are subject to swings in the economy. It is probably unwise to combine more than 2 or 3-years of data on such items as median income or the poverty rate.

## 2.6

### Combining Data From Several Surveys

A few items are included in more than one survey: health insurance is covered in NHIS, SIPP, and MEPS; other items are considered basic covariates for multivariate analysis in many surveys, but are important statistics in their own right. Age, sex, and marital status are almost defining characteristics, and they are collected in virtually all questionnaires. Other frequently obtained items are income (personal and/or family income), educational attainment, and labor force status. In Section 5 of this report we discuss the possibility of enhancing the subpopulations' sample sizes by combining data from several surveys.

One would like the question wordings to be reasonably consistent among the surveys that will be combined. This is probably not an issue for such demographic items as age, sex, and marital status, or for educational attainment. However, reporting of income, poverty and, to some extent, labor force status and occupation can be quite sensitive to both the question wording and the amount and type of probing carried out by interviewers. A major consideration for income, and possibly labor force, is how much discrepancy in question wording can be tolerated in order to provide a sufficient sample size for reasonable reliability. It may be possible to calibrate the results of various surveys so that adjusted data are in closer conformity.

One additional issue relating to comparability among surveys involves the population covered by the survey, that is: whether the samples represent all 50 states and D.C.; whether each survey includes the entire civilian non-institutional population, excludes some components, or includes some others, such as the military or institutional population. We do not expect this to be an important concern for most purposes, but analysts who are trying to establish historical series may find that even small inconsistencies can raise fundamental questions about the validity of the data.

## 2.7

### Response Rate and Other Quality Issues

The surveys listed in the Task 2 report, and which will be analyzed further in the balance of this report, are conducted by federal agencies, or carried out under contract for the agencies. Knowledgeable statistical staffs monitor the survey operations and the quality of the work is generally quite high. There is particular stress on attaining high response rates, and we believe that all of the surveys do about as well as can be expected, given that, with the exception

of the vital statistics system, the ACS, and the U.S. Census, the survey responses are voluntary rather than mandatory.

To say that the response rates are acceptable does not mean there are no potential nonresponse problems. All of the major surveys use poststratification in the final stage of weighting to reduce sampling errors, and to compensate as much as possible for nonresponse and undercoverage. There are almost always separate poststratification cells for blacks, Hispanics, and all other race/ethnic groups and NHANES has such cells for Mexican-Americans. The minority subgroups are almost always combined into categories like "total Hispanics" or "total other races" (which includes American Indians and Alaska Natives). Subdomains such as Puerto-Ricans, Cuban-Americans, Central-Americans, etc., are thus combined into a single class, with identical weights. Similarly, all Asians and Pacific Islanders get identical weights. If, in fact, some of these subgroups have lower response rates than the overall rate for the race/ethnic class, and are not separately adjusted, they will be underrepresented in the statistics. A similar situation exists with undercoverage. For example, if illegal aliens tend to avoid reporting (as seems likely) and if a higher proportion of Mexican-Americans are here illegally than in other Hispanic subpopulations (as is also likely), then the uniform weighting will slightly understate Mexican-Americans and overstate other Hispanic subgroups.

Unfortunately, there is not much that can be done to adjust for such occurrences. Agencies already make strenuous efforts to attain high response rates and it is unlikely that further exhortation to improve will be effective. Users of the data, however, should be aware of such limitations in drawing conclusions from the statistics.

## 2.8        **Estimation of Sampling Errors**

All of the major statistical agencies estimate sampling errors for their surveys. Since many of the surveys use complex, multi-stage sample designs, estimation of sampling errors is also fairly complex. Even further complications would result from the procedures for enhancing the sample as discussed above; averaging over a number of years, or combining the data from several surveys. Fortunately, software exists for the production of estimates of sampling errors for complex designs that could be adapted to cover averaging over years or combinations of surveys. We therefore see no reason to treat the ability to compute estimates of sampling errors as an impediment to the use of these procedures.

### 3. ABILITY OF CURRENT SURVEYS TO PROVIDE DATA WITH ADEQUATE PRECISION

#### 3.1 Standards for Precision

Sections 2.1 and 2.2 discussed ways of focusing on the precision levels required for various analytic uses. Since most U.S. Government surveys cover a broad array of data items, it is clear that no single standard of precision is likely to satisfy all potential uses of the data and that some compromises are necessary. It is particularly difficult to create standards of precision for a group of unrelated surveys whose specific analyses are yet to be developed at some future time. Under these circumstances, it seems sensible to use the standards for "generic" prevalence rates that were established for the study of the feasibility of producing state data from the NHIS. However, we reiterate our caveat that the standards may not satisfy all requirements, and if some new and critically important needs for statistical information on the subpopulations arise, the standards should be reviewed.

The precision levels that were examined for NHIS state level generic estimates were described in Section 2.2. We repeat them below:

- To achieve specific levels of precision for five "generic" prevalence rates: 0.01, 0.05, 0.10, 0.15, and 0.20.
- Three levels of precision are specified for the statistics described above: coefficients of variation (CV) of 30 percent, 20 percent, and 10 percent.

In order to remove these prevalence rates from the abstract, we show some examples reported in recent U.S. Government sponsored surveys:

Percent of U.S. population age 15 and over with earnings under \$10,000 in 1996	24.9% <sup>1</sup>
U.S. poverty rate in 1998	12.7% <sup>2</sup>
Percent of U.S. population without health insurance coverage during all of 1998	16.3% <sup>3</sup>
Percent of persons of Hispanic origin without health insurance in 1998	35.3% <sup>3</sup>
Cocaine use by adults employed full time	0.7% <sup>4</sup>
Cocaine use by adults employed part time	0.9% <sup>4</sup>
Cocaine use by unemployed adults	2.4% <sup>4</sup>

<sup>1</sup> Source: March 1997 CPS; U.S. Census Bureau Report P60, No. 206.

<sup>2</sup> Source: March 1999 CPS; U.S. Census Bureau Report P60, No. 207.

<sup>3</sup> Source: March 1999 CPS; U.S. Census Bureau Report P60, No. 208.

<sup>4</sup> Source: National Household Survey of Drug Abuse, 1997.

It may be useful to convert the generic coefficients of variation to standard errors and confidence intervals for a clearer view of the effects of the sampling errors on the statistics. They are shown below.

CV and prevalence rate	Standard error	66% confidence interval	95% confidence interval
<b>30% CV</b>			
.01	.003	.007-.013	.004-.016
.05	.015	.035-.065	.020-.080
.10	.030	.070-.130	.040-.160
.15	.045	.105-.195	.060-.240
.20	.060	.140-.260	.080-.320
<b>20% CV</b>			
.01	.002	.008-.012	.006-.014
.05	.010	.040-.060	.030-.070
.10	.020	.080-.120	.060-.140
.15	.030	.120-.180	.090-.210
.20	.040	.160-.240	.120-.280
<b>10% CV</b>			
.01	.001	.009-.011	.008-.012
.05	.005	.045-.055	.040-.060
.10	.010	.090-.110	.080-.120
.15	.015	.135-.165	.120-.180
.20	.020	.180-.220	.160-.240

Section 2.2 also notes the types of variables that are likely to have design effects beyond the average values used in this report. We have left them out of our discussion because we use a single average design effect (or, in a few cases, two) for each survey covered in this report.

Our examination assumes the prevalence rates are applied to the total of all persons in the subpopulation. In practice, analyses are frequently desired for subsets, e.g., adults or children, each sex separately, families rather than persons, persons below the poverty level, etc. Examining all possible uses of data would lead to such a wide variety of possibilities that no clear-cut decision could be made, and it seems sensible to restrict the alternatives. Basically, if subset analysis is considered of crucial importance for a survey, the sample size implied by each precision level can be thought of as applying to the subset, and the implications for the total sample for the survey can be calculated. For example, if a 30 percent CV requires a sample of 200 persons, then a sample of 400 persons is necessary if the same CV is desired for males and females separately.

Some examples of subsets, and associated prevalence rates are shown below.

Males 15 years and over: Percent with income under \$10,000	17.2% <sup>1</sup>
Hispanic males 15 years and over: Percent with income under \$10,000	21.4% <sup>1</sup>
Males, 15-24 years of age: Percent with income under \$10,000	40.8% <sup>1</sup>
Hispanic low-income persons (under 125% of poverty): percent without health insurance	44.0% <sup>2</sup>
Hispanic children under 18 years of age: percent without health insurance	30.0% <sup>2</sup>
Persons 65 years and over: percent without health insurance	1.1% <sup>2</sup>
Hispanic males, 65 years and over: percent with income below 50 percent of poverty	4.7% <sup>3</sup>

<sup>1</sup>Source: U.S. Census Bureau Report P60 No. 206.

<sup>2</sup>Source: U.S. Census Bureau Report P60 No. 208.

<sup>3</sup>Source: U.S. Census Bureau Report P60 No 207.

### 3.2 Effective Sample Sizes Required to Meet Precision Levels

The effective sample sizes needed to provide the precision levels for the various prevalence rates are shown in Table 3-1. They were derived from the simple formula for simple random sampling,

$$n = \frac{1-p}{p(CV)^2};$$

where  $p$  is the prevalence rate.

Table 3-1. Effective sample size needed for alternative prevalence rates and levels of precision

Prevalence rate	Precision level (CV)		
	.30	.20	.10
0.01	1,100	2,475	9,900
0.05	211	475	1,900
0.10	100	225	900
0.15	63	142	567
0.20	44	100	400

### **3.3 Survey Design Effects**

Section 2.4 of this report briefly described features of sample designs that affect the sampling errors and thus contribute to design effects. It was also noted that design effects can differ among statistics gathered in a survey, sometimes dramatically. As a basis for decision-making, we have chosen to use an average design effect for each survey, one that is approximately midway between the high and low values. In a few cases, we have indicated an additional design effect that applies to specific race/ethnic groups. However, an analyst who is concerned with a specific subject in a survey might prefer to use a different design effect that is more appropriate to the items to be studied. This report cannot take into account all possible analyses that could be carried out. We have tried to include enough information to permit modifications of the results for special subpopulations or items.

The design effects shown in Table 3-2 come from a number of sources (see Appendix A). Wherever possible, we have used published reports. These reports usually do not show design effects, as such, but the information on sampling errors makes it possible to calculate average design effects. When published reports with the required information were not available, the design effects were estimated from the descriptions of the sample design or through discussions with statisticians at the agencies sponsoring the surveys.

### **3.4 Nominal and Effective Sample Sizes**

Tables 3-3 to 3-5 show the sample sizes for all of the race/ethnic subgroups and are the same numbers reported in Tables A-1 to A-3 of the Task 2 report. As noted, these data represent approximations of the number of sample cases for each subpopulation, and were obtained either from published reports of the Federal agencies sponsoring the survey, provided by the agencies, or derived by Westat. We refer to these numbers as the “nominal sample sizes” to distinguish them from the effective sample sizes. We note that we have included all race/ethnic subgroups, including those that are not currently identified in the data set. The sources used to provide estimates of design effects are shown in Appendix A.

Table 3-2. Average design effects<sup>1</sup> for minorities

Survey	Average design effect
Census	
Census 2000	1.0
ACS	1.0
CPS <sup>2</sup>	1.5
SIPP – Hispanic	2.4
SIPP – API and American Indians or Alaska Natives	1.6
NCHS/CDC	
NHIS – Hispanics and American Indians or Alaska Natives	1.5
NHIS – API	1.3
NSFG – Hispanics and American Indians or Alaska Natives	1.7
NSFG – API	1.4
NIS	1.3
NHANES <sup>3</sup> – Mexican-American	2.2
NHANES <sup>3</sup> – Other minorities	1.8
AHRQ	
MEPS – Hispanic	1.2
MEPS – API and American Indians or Alaska Natives	2.1
HCFA	
MCBS	1.1
SAMHSA	
NHSDA	2.2
NCES	
NHES <sup>4</sup>	1.4
ECLS-B	1.2
ECLS-K <sup>5</sup>	2.5

<sup>1</sup> Most of the surveys are based on household samples. The design effects apply to statistics that do not cluster strongly within households, e.g., health conditions, educational attainment, and labor force status. Items like poverty status, availability of health insurance, urban-rural residence, etc. generally are identical for all members of a household, and the design effects for such items are much larger, usually two to three times the ones shown in the table.

<sup>2</sup> The design effects are approximately the same for the March CPS and for other months.

<sup>3</sup> The design effects are those of statistics on data for the total of each race/ethnic group. Design effects for individual age-sex groups are lower.

<sup>4</sup> The design effects shown apply to statistics on children who constitute the main focus of NHES. Data for adults are sometimes included in the survey, and they are subject to higher design effects.

<sup>5</sup> The design effect shown, 2.5, applies to most social, economic, and related items. The design effect for test scores is about 5.

Many of the U.S. Government surveys are repetitive, that is either carried out every year, conducted several times a year, or as in the case of CPS, conducted every month. In most cases, the sample sizes shown in this report describe the annual sample as it was in the time period noted. The reader should be aware that sample sizes are sometimes changed because of budgetary restrictions or other causes. For analysis of a data set, it would be useful to ascertain whether there is an important difference in the sample design between the time period analyzed and the reference date shown in Section 1.2. If so, the sample sizes should be modified accordingly. There are a few cases in which there may be some ambiguity in the sample size. A brief discussion of these cases follows:

- **CPS.** The CPS is carried out monthly, primarily to obtain labor force information. In March of each year, the CPS becomes a mini-census, including such information as income, mobility, family and household composition, and related data. Supplementary items are also covered in other months, such as school enrollment, children ever born and voting (in alternate years.) The sample sizes are virtually identical in 11 of the 12 months. In March, the number of Hispanics in the sample is doubled, with non-Hispanics kept the same as in other months.

The sample size in each of the 11 months (excluding March) is referred to as the CPS-Monthly sample. The March sample is similarly referred to as CPS-March. In analyzing a CPS data set for Hispanic subgroups, it therefore is important to identify the month in which the information was obtained. The March sample sizes for Asian and Pacific Islanders and American Indians or Alaska Natives are the same as in other months, so that month of data collection does not affect the sample size.

Since labor force data are collected each month in CPS, it is possible to obtain yearly averages by pooling the data sets for all 12 months of a year. However, the CPS sample retains the same sample units in a 4-month cycle, and there is about a 75 percent overlap in the sample from one month to the next. The effectiveness of the annual sample is thus very much less than 12 times the monthly sample. Section 4.5 of this report discusses the sampling errors of annual averages in CPS.

- **SIPP.** The terms “panels” and “waves” are used to describe the SIPP sample. Panels refer to the set of households that comprise a probability sample of the total population and also of subpopulations. Waves refer to the interview cycles; each panel is interviewed several times a year (i.e., several waves) and over the course of a number of consecutive years. The waves reflect the fact that SIPP is mainly viewed as a source of longitudinal data. A panel’s sample size over the course of the interview waves is intended to be the same, although there is normally some attrition resulting from cumulative nonresponse. This moderate attrition does not change the sample sizes sufficiently to affect any conclusions in this report.

Currently, a single panel is used. The current panel was introduced in 1996 and will continue through 1999. The sample sizes for SIPP shown in this report are those of the current panel. In earlier years, a rotating panel structure was used, with several panels operating in each year. Before proceeding with a study based on SIPP, an analyst should check the number and size of panels used in the time period of interest. It should also be noted that with the current lack of rotation of the panel, there is very

little to be gained by combining years (except, of course, for longitudinal analyses of changes over time.)

- **NHANES.** Currently, each year, a new sample of about 5,000 individuals of all ages (comprised of about 1,500 Mexican-Americans, 82 "Other Hispanics," 113 Asian and Pacific Islanders and 24 American Indians or Alaska Natives) is interviewed and examined. The samples among years are independent; consequently the results can be aggregated across years to improve reliability. Most of the past analyses of NHANES have concentrated on the health and nutrition status of detailed age-sex-race/ethnicity groups, and a 6-year accumulation of data is necessary to meet the established precision requirements for these detailed age groups. Shorter periods can be used for broader age groups, and NHANES III used both a 3-year and 6-year accumulation. The use of independent samples each year in the current NHANES permits considerable latitude for the analyst in combining years and it is expected that most analyses will use several years.

The sample sizes shown in this report refer to a single year's sample. Section 4.4 discusses the effect of combining years of data.

- **MEPS and MCBS.** Each year's sample in these two surveys is interviewed several times during the course of the year. The purpose of the multiple visits is to shorten the time period for which information is obtained and, thus, reduce the possibility that memory factors will affect the quality of data. The data tapes combine data obtained in the multiple visits, and the sample sizes shown refer to the number of persons for whom annual data are obtained, not the number of interviews.

Currently, each year's MEPS sample consists of two panels; one introduced for the first time that year, and the second carried over from the preceding year. For this reason it is important that sample sizes be verified before attempting to utilize these data. The MEPS sample sizes shown in this report refer both to the new panel introduced in 1999 and the panel carried over from 1998. The MCBS also has a panel structure and we report the sample size for the four panels included in early 1998.

- **NVS.** All births and deaths are covered in the NVS Mortality and Natality data sets. As indicated in Section 2.1, this report focuses on the uses of the data for descriptive analyses. From this viewpoint, vital statistics are not subject to sampling error. Consequently, the two NVS data sets are not included in either the tables on effective sample sizes or the subsequent discussion of available precision.

Table 3-3. Approximations of Hispanic sample cases in the data set

Data set	Total Hispanic	Mexican-Americans	Puerto Ricans	Cubans	Central or South American	Other Hispanic
<b>Census</b>						
Census 2000 <sup>1</sup>	4,508,000	2,850,000	475,000	190,000	650,000	335,000
ACS	900,000	570,000	95,000	38,000	130,000	67,000
CPS-March	11,260	6,940	1,190	470	1,685	975
CPS-Monthly	5,635	3,470	595	235	845	490
SIPP	10,845	7,181	1,172	372	1,306	814
<b>NCHS/CDC</b>						
NHIS	22,145	13,869	2,353	1,165	2,093	4,758
NSFG	2,097	1,330	221	88	302	156
NIS	4,852	3,529	398	99	526	300
NHANES	1,582	1,500	24	10	32	16
<b>AHRQ</b>						
MEPS	5,375	3,650	600	225	766	134
<b>HCFA</b>						
MCBS	464	254	42	67	52	50
<b>SAMHSA</b>						
NHSDA	5,000	3,170	527	211	721	372
<b>NCES</b>						
NHES	18,804	13,675	1,541	385	2,040	1,162
ECLS-B	1,979	1,367	160	35	137	280
ECLS-K	2,957	2,150	242	61	321	183

<sup>1</sup> Long form data

NOTE: The sample cases for each data set reflect the population coverage of the respective surveys. For example, CPS-March covers all persons in the civilian noninstitutional population, whereas NSFG covers women 15 to 44 years of age. The Task 2 descriptions of the respective data sets note the appropriate population coverage. The sample sizes are the number of sample persons in each subgroup, including those that are not identified in the data file.

Table 3-4. Approximations of Asian and Pacific Islander sample cases in the data set

Data set	Total Asian and PI	Chinese	Filipinos	Japanese	Asian Indian	Korean	Vietnamese	Hawaiian	Other
Census									
Census 2000 <sup>1</sup>	1,580,000	375,000	300,000	180,000	175,000	175,000	135,000	45,000	195,000
ACS	316,000	75,000	60,000	36,000	35,000	35,000	27,000	9,000	39,000
CPS-March	4,555	995	850	515	495	485	375	125	565
CPS-Monthly	4,555	995	850	515	495	485	375	125	565
SIPP	3,293	745	637	386	370	362	280	95	421
NCHS/CDC									
NHIS	3,284	755	647	356	320	342	356	112	396
NSFG	327	74	63	38	37	36	28	9	42
NIS	1,172	265	227	137	131	129	100	33	150
NHANES	113	27	22	13	12	12	10	3	14
AHRQ									
MEPS	750	152	170	62	111	96	45	17	97
HCFA									
MCBS	151	34	29	18	17	17	13	4	19
SAMHSA									
NHSDA	700	158	135	82	78	77	59	20	90
NCES									
NHES	4,420	999	855	517	495	486	376	128	566
ECLS-B	2,483	705	467	134	282	278	217	74	325
ECLS-K	1,870	423	362	219	209	206	159	54	239

<sup>1</sup> Long form data

NOTE: The sample cases for each data set reflect the population coverage of the respective surveys. For example, CPS-March covers all persons in the civilian noninstitutional population, whereas NSFG covers women 15 to 44 years of age. The Task 2 descriptions of the respective data sets note the appropriate population coverage. The sample sizes are the number of sample persons in each subgroup, including those that are not identified in the data file.

Table 3-5. Approximations of American Indian or Alaska Native sample cases in the data set

Data set	American Indian and Alaska Native
Census	
Census 2000 <sup>1</sup>	330,000
ACS	67,000
CPS-March	1,600
CPS-Monthly	1,350
SIPP	1,200
NCHS/CDC	
NHIS	978
NSFG	77
NIS	460
NHANES	24
AHRQ	
MEPS	375
HCFA	
MCBS	25
SAMHSA	
NHSDA	166
NCES	
NHES	1,675
ECLS-B	50
ECLS-K	364

<sup>1</sup> Long form data

NOTE: The sample cases for each data set reflect the population coverage of the respective surveys. For example, CPS-March covers all persons in the civilian noninstitutional population, whereas NSFG covers women 15 to 44 years of age. The Task 2 descriptions of the respective data sets note the appropriate population coverage. The sample sizes are the number of sample persons in each subgroup, including those that are not identified in the data file.

The effective sample sizes are simply the nominal sample sizes divided by the design effects. They are shown in Tables 3-6 to 3-8. The effective sample sizes will be used to identify data sets that satisfy minimum standards of reliability.

Table 3-6. Effective sample sizes for Hispanics

Data set	Total Hispanic	Mexican-American	Puerto Rican	Cuban	Central or South American	Other Hispanic
<b>Census</b>						
Census 2000 <sup>1</sup>	4,508,000	2,850,000	475,000	190,000	650,000	335,000
ACS	900,000	570,000	95,000	38,000	130,000	67,000
CPS-March	7,507	4,627	793	313	1,123	650
CPS-monthly	3,757	2,313	397	157	563	327
SIPP	4,519	2,992	488	155	544	339
<b>NCHS/CDC</b>						
NHIS	14,763	9,246	1,569	777	2,093	1,079
NSFG	1,234	782	130	52	178	92
NIS	3,732	2,715	306	76	405	231
NHANES	727	682	12	6	18	9
<b>AHRQ</b>						
MEPS	4,479	3,042	500	188	637	112
<b>HCFA</b>						
MCBS	422	231	38	61	47	45
<b>SAMHSA</b>						
NHSDA	2,273	1,441	240	96	328	169
<b>NCES</b>						
NHES	13,431	9,768	1,101	275	1,457	266
ECLS-B	1,649	1,139	133	29	114	233
ECLS-K	1,183	860	97	24	128	73

<sup>1</sup> Long form data

NOTE: The sample cases for each data set reflect the population coverage of the respective surveys. For example, CPS-March covers all persons in the civilian noninstitutional population, whereas NSFG covers women 15 to 44 years of age. The Task 2 descriptions of the respective data sets note the appropriate population coverage. The sample sizes are the number of sample persons in each subgroup, including those that are not identified in the data file.

Table 3-7. Effective sample sizes for API

Data set	Total API	Chinese	Filipino	Japanese	Asian Indian	Korean	Vietnamese	Hawaiian	Other
Census									
Census 2000 <sup>1</sup>	1,580,000	375,000	300,000	180,000	175,000	175,000	135,000	45,000	195,000
ACS	316,000	75,000	60,000	36,000	35,000	35,000	27,000	9,000	39,000
CPS-March	3,037	663	567	343	330	323	250	83	377
CPS-monthly	3,037	663	567	343	330	323	250	83	377
SIPP	2,058	466	398	241	231	226	175	59	263
NCHS/CDC									
NHIS	2,433	559	479	264	237	253	264	83	293
NSFG	234	53	45	27	26	26	20	6	30
NIS	902	204	175	105	101	99	77	26	115
NHANES	63	15	12	7	7	7	5	2	8
AHRQ									
MEPS	357	72	81	30	53	46	21	8	46
HCFA									
MCBS	137	31	26	16	15	15	12	4	17
SAMHSA									
NHSDA	318	72	61	37	35	35	27	9	41
NCES									
NHES	3,157	714	611	369	354	347	269	91	404
ECLS-B	2,069	588	389	112	235	232	181	62	271
ECLS-K	748	169	145	88	84	82	64	22	96

<sup>1</sup> Long form data

NOTE: The sample cases for each data set reflect the population coverage of the respective surveys. For example, CPS-March covers all persons in the civilian noninstitutional population, whereas NSFG covers women 15 to 44 years of age. The Task 2 descriptions of the respective data sets note the appropriate population coverage. The sample sizes are the number of sample persons in each subgroup, including those that are not identified in the data file.

Table 3-8. Effective sample size for American Indians or Alaska Natives

Data set	Effective sample size
<b>Census</b>	
Census 2000 <sup>1</sup>	330,000
ACS	67,000
CPS-March	1,067
CPS-monthly	1,067
SIPP	1,000
<b>NCHS/CDC</b>	
NHIS	652
NSFG	45
NIS	354
NHANES	12
<b>AHRQ</b>	
MEPS	179
<b>HCFA</b>	
MCBS	23
<b>SAMHSA</b>	
NHSDA	75
<b>NCES</b>	
NHES	1,196
ECLS-B	148
ECLS-K	146

<sup>1</sup> Long form data

NOTE: The sample cases for each data set reflect the population coverage of the respective surveys. For example, CPS-March covers all persons in the civilian noninstitutional population, whereas NSFG covers women 15 to 44 years of age. The Task 2 descriptions of the respective data sets note the appropriate population coverage. The sample sizes are the number of sample persons in each subgroup, including those that are not identified in the data file.

### 3.5

### Surveys and Race/Ethnicity Groups Meeting Standards for Precision

A comparison of the effective sample sizes in Tables 3-6 to 3-8 with the numbers needed to meet alternate levels of precision shown in Table 3-1 indicate which race/ethnic subgroups meet these standards for each of the surveys.

We should like to reiterate the caveats mentioned earlier in the discussion of these standards. The sample sizes in Table 3-1 will provide the coefficient of variation for the indicated estimate of prevalence of the total population in the race/ethnic subgroup (or of the total target population of the survey; e.g., females 15-44 for NSFG, person 65 or older for MCBS, etc.) If the contemplated analysis includes examining subsets of the total, such as individual age groups, urban-rural residence, or low-income vs. higher-income persons, much larger sample sizes are needed; essentially each subset would require approximately the sample sizes shown in Table 3-1. Since the specific studies to be carried out have not yet been developed, this report does not contain a provision for subset analysis, but the possibility of the need for such statistical breakdowns and their implications should be kept in mind.

Most of the surveys use an identical sampling rate for all persons in each race/ethnic group. In these surveys, the sample size for any subset can be estimated by taking the proportion of the sample equal to the proportion of the relevant population in that subset. For example, for analysis of data by gender, the male (and female) sample will be equal to about one-half the total sample. Similarly, for an age group containing about 20 percent of the relevant population, the sample will be 20 percent of the total sample in the race/ethnic subgroup. Similar relationships hold for other subsets, such as regional breakdowns, income classes, etc. For subset analyses, the nominal and effective sample sizes in the tables, which follow, should be adjusted to reflect the portion of the subgroup to be analyzed.

There are a few exceptions to the use of a common sampling rate for all members of a subgroup. NHANES focuses on 52 age-sex-race/ethnicity subsets, and uses approximately the same sample sizes for each. The 52 groups are described in several reports on the methodology of NHANES, and analysts concerned with subsets of the race/ethnicity subgroups should refer to the NHANES publications for appropriate methods of estimating the sample sizes. SIPP oversamples persons in poverty. For subset analyses comprising persons in poverty (or items correlated with poverty), the analyst should obtain a description of the current SIPP sample and use it to estimate the sample size.

Secondly, the design effects in Table 3-2 that were inputs to the calculation of the effective sample sizes basically apply to data that are not heavily clustered within households. Examples of statistics that are not clustered, or only moderately clustered are: smoking status, presence of specific

chronic illnesses such as hypertension or arthritis, occupation, and very large expenditures for medical care during the year. For such items, members of a household are unlikely to have the same characteristics. On the other hand, as is indicated in footnote 1 of Table 3-2, items such as poverty status, health insurance, urban-rural residence, etc. tend to be identical for all members in a household, and the design effects are usually two to three times as large as those in Table 3-2. Other examples of items with high clustering effects are: mobility status, whether or not foreign born, and income class. Such items will tend to be identically reported within a household so that obtaining the statistics from all members of a household is no more useful than an interview with only one household member. In such instances, the design effect is increased by a factor equal to the average household size, that is by a factor of about 3.5 for Asian and Pacific Islanders, 4.3 for American Indians and Alaska Natives and 3.6 for Hispanics. The average household size (and consequently the design effects) can differ among the subgroups that are the focus of this report. For example, the average household size for Hispanic subgroups varies from a low of 2.6 for Cubans to 3.9 for Mexicans. An analyst should check the household sizes of the subgroups to be studied if highly clustered items are important variables, and modify the design effects accordingly. An alternate way of accomplishing the same goal for highly clustered items is to treat the sample size as the number of households in the sample rather than the number of persons. The nominal and effective sample sizes in the various tables should then be divided by the average household size.

Results of the comparisons of Tables 3-6 to 3-8 with Table 3-1 are summarized below. Table 3-1 indicates the sample size cut-offs for various levels of confidence in the data. Thus, an effective sample size of 500 satisfies requirements for a 20 percent CV for all prevalence rates except very rare ones (i.e.,  $p = .01$ ); an effective sample of 1,000 will provide a CV of .10 on prevalence rates greater than or equal to .10, as well as satisfying the criteria mentioned for a sample of 500; and a sample size of about 2,000 to 2,500 will produce CVs of .20 or better for prevalence rates as low as .01.

- **Census 2000/ACS.** It is clear that the Census 2000 and the ACS samples are sufficiently large to satisfy any reasonable precision requirement.
- **CPS.** Both the March CPS, which oversamples Hispanics by a factor of 2, and the monthly CPS satisfy virtually all requirements for Mexican-Americans except the most stringent one, i.e., the CV of .10 on a prevalence rate of .01. (The monthly sample is a little short of what is needed for a CV of .20 on a rate of .01, but the difference is negligible.)

For prevalence levels of .05 or greater, the March sample of Central or South Americans satisfies almost all of the requirements. The March samples of Puerto Ricans and “other” Hispanics is satisfactory for prevalence rate of .05 or greater when CVs of .20 or more are required, and for rates of .15 or greater when a CV of .10 is needed.

Other than for March, the monthly CPS samples for all Hispanic groups except Mexican-Americans and Central or South Americans are fairly small and only provide the sample needed for CVs of .20 or greater with rates of .05 or more. This is also true of the March Cuban sample; the monthly Cuban sample produces even less reliability.

The CPS American Indian or Alaska Native sample is sufficient for CVs of .20 or greater with prevalence rates of .05 or greater. It is also large enough to provide a CV of .10 when the prevalence rates are .10 or greater.

The Chinese and Filipino samples are inadequate when the rate is as low as .01, but will provide CVs of .20 for rates .05 or greater, and CVs of .10 for rates above .15. The Japanese, Asian-Indian, Korean and "other" API samples are quite similar and are mostly sufficient to provide CVs of .20 for rates of .10 or greater and a CV of .30 for a .05 rate. The Hawaiian sample is quite small and satisfies hardly any of the requirements.

- **SIPP.** The Wave 1, 1996 Mexican-American sample meets all the precision requirements except the most stringent one that is to achieve a 10 percent CV on a .01 prevalence rate. The Puerto Rican and Central and South American sample will achieve a CV of .20 or better for prevalence rates of .05 or greater. Estimates for Cubans and Other Hispanics only will fulfill rather modest requirements. The Chinese, Filipino and American Indian or Alaska Native samples satisfy moderate requirements, but the Hawaiian sample is quite small and can only satisfy the most generous requirements. The effectiveness of the sample may weaken somewhat for successive interviewing waves, as cumulative nonresponse affects the sample size.
- **NHIS.** As a result of the oversampling of Hispanics, the annual Mexican-American sample fulfills the requirements for all CV and prevalence rates, except for a CV of .10 on a .01 prevalence rate, and it comes close to meeting that goal.

The Central and South American annual sample satisfies the criteria for precision for prevalence rates of .05 or greater. The Puerto Rican and "other" Hispanic annual samples meet all the requirements when the prevalence rate is .05 or greater, except the goal of a CV of .10 for a prevalence rate of .05. The smaller Cuban sample is still large enough to obtain a CV of .20 or better for prevalence rates of .05 or greater, and to provide a CV of .10 for prevalence rate of .15 or more.

The annual American Indian or Alaska Native sample is close to that of Cubans, and will achieve approximately the same levels of precision.

The Chinese and Filipino samples are a little smaller than the American Indian or Alaska Native sample, but they still will satisfy similar goals, that is, they will provide CVs of .20 or better for prevalence rates of .05 or greater. The other Asian and Pacific Islander groups will only meet the most modest criteria, a .30 CV for rates of .05 or greater and a .20 CV when the rate is .10 or more.

- **NSFG.** Mexican-Americans comprise the only population subgroup that can provide reasonable precision—a CV of .20 for prevalence rates of .05 or more and a CV of .10 for rates of .15 or more. All of the other subgroups could satisfy only very minimal standards.

- **NIS.** The Mexican-American sample satisfies all precision requirements except for a CV of .10 on a .01 prevalence rate. None of the other race/ethnic subgroups do very well. The Puerto Rican, Central and South American, and American Indian or Alaska Native samples can meet moderate standards—a .20 CV on prevalence rates of .10 or greater and a .30 CV on a prevalence rate of .05. The other race/ethnic subgroup could provide only crude estimates.
- **NHANES.** The NHANES sample was designed specifically to provide good reliability for Mexican-Americans, but only when several years of data are combined. The annual sample size is fairly modest and will only provide a CV of .20 for prevalence rates equal to or greater than .20. A CV of .10 is achieved for prevalence rates equal to or greater than .15. None of the other race/ethnic groups can provide usable annual data.
- **MEPS.** The Mexican-American sample satisfies all of the precision requirements for prevalence rates of .05 and greater and even does fairly well with rates of .01. Some modest analysis is possible for Puerto Ricans and Central or South Americans. The samples on the other race/ethnic subgroups are too small to be useful.
- **MCBS.** The sample sizes are too small for subgroup analyses, even for Mexican-Americans.
- **NHSDA.** Mexican-Americans could provide a CV of .10 for prevalence rates of .10 or more and CVs of .20 for rates of .05. Some limited analysis is possible of Puerto-Ricans and Central or South Americans. None of the other subgroups would provide useful data.
- **NHES.** The Mexican-American sample meets, or comes very close to meeting, all of the precision requirements. The Puerto-Rican, Central or South American, and American Indian or Alaska Native samples are reasonably large, and would produce a CV of .20 for prevalence rates of .05, and a CV of .10 for rates of .10 or greater. The Chinese and Filipino samples are large enough for a CV of .20 with a prevalence rate of .05 or greater. Only limited use is possible of the other population subgroups.
- **ECLS-B.** Using the sample sizes at the initial interview, Mexican-Americans will provide a CV of .20 or better for prevalence rates of .05 or more, and a CV of .10 or better on rates of .10 or greater. The Chinese sample will achieve CVs of .20 or better on rates of .05 or greater. The other subgroups would satisfy only very minimal requirements.
- **ECLS-K.** Subgroup analysis of the results of the Year 1 interview essentially would have to be restricted to Mexican-Americans. Their estimates of prevalence rates of .10 or greater would be subject to a CV no greater than .10, and prevalence rates of .05 would have a CV of .20.

The analysis above can be summarized as follows. The vital statistics records, Census 2000 and the ACS will permit detailed and complex analyses of all race/ethnic subpopulations. The March CPS, the NHIS, and NHES can produce quite accurate statistics for Mexican-Americans, moderately good data for Puerto-Ricans and Central or South Americans, and acceptable data for the other Hispanic subgroups, with the possible exception of Cubans. Data for Chinese, Filipinos, and American Indian or

Alaska Native would be fairly reliable. Only limited analysis could be made of data for the remaining API subgroups. The monthly CPS and SIPP would be weaker for Hispanics, but mostly still acceptable. For the other surveys, acceptable precision is only possible for Mexican-Americans, and MCBS would not even be acceptable for that subgroup.

It is important to remember that the above analyses apply to the ability of the surveys to provide acceptable accuracy on prevalence rates (or percentage distributions) of total persons in each subpopulation. Many surveys require examination of important subsets of the population, as well as the total. For example, NHANES concentrates on age-sex-race/ethnicity subgroups, MEPS examines low-income persons as well as the total population, and an analytic group in the NSFG is teenagers, by race/ethnicity. For such analyses, the survey needs to have each subset have the sample sizes in Table 3-1. Thus, a simple four-way breakdown of the population, such as persons under or over 25 years by sex, would require a sample four times as great as the numbers in Table 3-1.

Table 3-9 contains guidance on the ability of the various databases to provide acceptable precision levels, as follows:

- A. Detailed cross-classification is possible with reasonable precision;
- B. Some limited cross-classification is possible;
- C. Only simple distributions are possible; and
- D. No analysis is possible.

The classifications are subjective, and it is possible to reach different conclusions on the levels of precisions that are reasonable. An analyst should determine how much error can be tolerated before reaching a conclusion on the detailed analysis to be carried out. Once again, given the possible changes in sample size or design, as well as the use of overlapping samples, we urge that, prior to using a particular data file, the current sample sizes and design effects be verified.

Table 3-9. Adequacy of databases for provision of data with acceptable precision (see footnote\* for description of codes used)

Database	Hispanic					American Indian or Alaska Native
	Mexican-American	Puerto Rican	Cuban	Central & South American	Other	
Census						
Census 2000	A	A	A	A	A	A
ACS	A	A	A	A	A	A
CPS-March	A	C	C	B	C	B
CPS-Monthly	B	C	D	C	C	B
SIPP	B	C	D	C	C	B
NCHS/CDC						
NHIS	A	B	C	B	B	C
NSFG	C	D	D	D	D	D
NIS	B	C	D	C	C	C
NHANES	C	D	D	D	D	D
AHRQ						
MEPS	B	C	D	C	C	D
HCFA						
MCBS	C	D	D	D	D	D
SAMHSA						
NHSDA	B	C	D	C	D	D
NCES						
NHES	A	B	C	B	C	B
ECLS-B	B	D	D	D	C	D
ECLS-K	C	D	D	D	D	D

\* Level of detail possible that can be attained with adequate precision

- A Detailed cross-classification possible
- B Some limited cross-classification
- C Only simple distributions
- D Analysis not possible

Effective sample sizes

- 4,000 or more
- 1,000 to 3,999
- 200 to 999
- Under 200

Table 3-9. Adequacy of databases for provision of data with acceptable precision (see footnote\* for description of codes used) (continued)

Data set	Chinese	Filipino	Japanese	Asian Indian	Korean	Vietnamese	Hawaiian	Other
<b>Census</b>								
Census 2000	A	A	A	A	A	A	A	A
ACS	A	A	A	A	A	A	A	A
CPS-March	C	C	C	C	C	C	D	C
CPS-Monthly	C	C	C	C	C	C	D	C
SIPP	C	C	C	C	C	D	D	C
<b>NCHS/CDC</b>								
NHIS	C	C	C	C	C	C	D	C
NSFG	D	D	D	D	D	D	D	D
NIS	C	D	D	D	D	D	D	D
NHANES	D	D	D	D	D	D	D	D
<b>AHRQ</b>								
MEPS	D	D	D	D	D	D	D	D
<b>HCFA</b>								
MCBS	D	D	D	D	D	D	D	D
<b>SAMHSA</b>								
NHSDA	D	D	D	D	D	D	D	D
<b>NCES</b>								
NHES	C	C	C	C	C	C	D	C
ECLS-B	C	C	D	C	C	D	D	C
ECLS-K	D	D	D	D	D	D	D	D

\* Level of detail possible that can be attained with adequate precision

- A Detailed cross-classification possible
- B Some limited cross-classification
- C Only simple distributions
- D Analysis not possible

Effective sample sizes

- 4,000 or more
- 1,000 to 3,999
- 200 to 999
- Under 200

The ability to produce acceptable data also depends on whether the survey collects the detailed race/ethnicity description of each sample person and enters the code in the data set. The Task 2 report indicated a few cases in which not all subpopulations were identified. Many of the surveys simply ask whether the sample person is an Asian or Pacific Islander without obtaining additional detail. The NVS, both natality and mortality, record the identification of Chinese, Japanese, Hawaiian, and Filipinos in all 50 states, but identify the other ethnic groups—Vietnamese, Asian-Indian, Korean, Samoans, and Guamanians—in only nine states which contain about two-thirds of the U.S. population in each of these groups. Obviously, the identifications and coding in the surveys and the NVS would need to be expanded to make tabulations possible.

## 4. COMBINING DATA FOR SEVERAL YEARS

It is clear that with the exception of the National Vital Statistics data sets, the Census 2000, and the ACS, the surveys can provide only limited information on race/ethnic subpopulations. The Mexican-American samples are adequate in most of the surveys but cross-classifications will rarely be possible for the other groups. Sections 4, 5, and 6 describe ways of enhancing the samples. In this section we discuss what is probably the simplest and least costly way of doing this, that is combining several years of data. The discussion, of course, omits the NVS, Census 2000, and the ACS, since the existing sample sizes are fully adequate.

### 4.1 Annual vs. Surveys Carried out at Periodic Intervals

Combining years of data is only practical for surveys that are carried out one or more times per year. Some of the surveys are conducted at periodic intervals. Although it would be possible to combine several cycles of such surveys, the length of time covered—probably 10 years or more—would make the results of doubtful utility. Also, SIPP uses the same households over a number of years, so that combinations of years do not provide much additional information.

The annual surveys for which combinations of years are practical are the CPS (March and monthly), NHIS, NHANES, NIS, MEPS, MCBS, and NHDSDA. NHES has been omitted since there is a different emphasis in subject matter each year, so that it falls closer to periodic than annual surveys.

The plans for current NHANES implicitly assume that the detailed analyses of the survey data will be based on averages over a number of years. Each year of current NHANES is based on a representative sample of about 5,000 persons in total, far too few to provide acceptable data for the many age-sex-race/ethnicity domains NCHS considers important to study. Combinations of years will be used for analyses of these domains, probably up to 6-years for the most detailed groups. In some ways, this can be considered a model for annual averages for other surveys.

### 4.2 Maximum Number of Years for Reasonable Analysis

Section 2.6 of this report pointed out that the maximum number of years for which combined data would be meaningful depended on the specific item. Most health related items and fertility patterns

change rather slowly over time, and the most recent 3 to 5-year averages will generally reflect current conditions reasonably well. In fact, the NHIS has published 3-year average data for Asian and Pacific Islanders (as a combined group), so a precedent exists. Economic statistics, however, are likely to be much more volatile; thus the time period should be considerably shorter. (However, in the absence of any other data, even somewhat outdated information such as a 3-year average, will be better than relying on the decennial census as the source of information for the full intercensal period. It is interesting to note that the ACS is planning to combine up to 5 years of data in order to produce reliable, small area data.)

To provide the greatest flexibility for users of this report, we will examine the improvement in precision for three combinations of years — 2, 3, and 5 years.

#### **4.3 Effective Sample Size for Combined Years**

The effective sample sizes for combined years are shown in Tables 4-1 to 4-3. It can be seen that except for NHANES, the effective sample sizes for 2 years are a little less than twice the sample for a single year; similarly the 3 and 5 year effective samples are not quite 3 or 5 times the annual sample sizes. All of the surveys use clustered sample designs and a sequence of several years' samples are mostly in the same clusters, or in neighboring ones. The lack of independence among several years' samples tends to reduce the effective sample size. We have estimated that the reduction in effective sample size over a 2-year interval is about 17 percent; the reduction for a 3-year period is 25 percent; and the reduction for 5 years is about 35 percent. These come from estimated year-to-year correlations in the sample: year-to-year correlations are expected to average .20, 2 years apart correlations are .10, 3 years apart are .07, and 4 years are .05. The current NHANES samples are independent across years, and, therefore, there is no reduction in effective sample size.

The effective sample sizes in Tables 4-1 to 4-3 are approximations based on even more assumptions and averages than the numbers in Tables 3-6 to 3-8. The sample sizes in each year are subject to sampling errors, and to the vagaries of erratic response rates. This is especially true for the minority subgroups with very small samples; the samples for Cubans or Hawaiian could differ in neighboring years by 10 or 20 percent from the year reflected in our tables. Also, the year-to-year correlations, resulting from the similarities in characteristics in neighboring households, are average values expected over a set of items, similar to the use of average design effects. Nevertheless, the numbers shown in Tables 4-1 to 4-3 indicate the order of magnitude of effective sample sizes and reveal whether useful analyses are possible from each of the data sets.

One feature of the monthly CPS sample should be noted. The monthly CPS includes two kinds of data sets: (1) labor force information and critical demographic items (e.g., age, sex, household relationship, etc.) obtained each month; and (2) supplemental items covered in months other than March. The supplemental items (based on the monthly CPS sample sizes) that are likely to be of greatest interest are number of children ever born, related fertility information, and school enrollment. Voting registration and behavior in the most recent election is obtained every second year, but it is doubtful that combining pairs of years would be meaningful. Voting patterns in presidential and non-presidential years are very different, and such combinations would probably not be analytically revealing. The entries for CPS in Tables 4-1 to 4-3 are restricted to the supplemental items. The annual sample sizes for labor force information, of course, are much larger than the numbers shown, since they are comprised of 12 monthly samples (see Section 4.5.) The supplemental items included in the March interview are based on the same sample as the other monthly supplements, except for Hispanics for whom the sample is doubled.

Since the MEPS and NSFG samples are taken from NHIS respondents, it would be possible to supplement their samples with additional names and addresses from the NHIS. These names and addresses would be a few years old, and thus it may be more convenient to simply combine multiple years of MEPS respondents. The exact timing of these surveys, and the associated costs, would have to be examined before a decision is made on which approach would be preferable.

#### **4.4 Surveys Meeting Standards for Precision**

Section 3.6 discussed the ability of the surveys to produce reasonable precision in the analysis of the subpopulations or for crossclassifications within these subpopulations. (“Reasonable precision” is based on a subjective judgment of the importance of meeting the various standards described earlier, i.e., CVs of 30 percent, 20 percent, and 10 percent for prevalence rates of .01, .05, .10, .15, and .20.) We will use the same criteria to evaluate the analytic ability of combinations of several years of survey data. As in the case of data for a single year, some studies may need greater precision and others less, and analysts should consider whether they need to modify the summary below.

Three or 5-year averages for CPS supplemental items collected in a single month for a given year would provide sample sizes large enough to satisfy analytic needs for most Hispanic subgroups, although only limited cross-classifications would be possible for Cuban-Americans. Fairly detailed analyses would be possible for American Indians or Alaska Natives, and for Chinese and Filipinos. Less detailed cross-classifications would be available for most of the other API subgroups, and only simple distributions of Hawaiian would have reasonable reliability.

The NHIS Hispanic sample is quite large, and a 2 or 3-year combination will provide quite reliable data, including cross-classifications, for all Hispanic subgroups, and moderately detailed cross-classifications for Cuban-Americans. A 5-year average will permit quite detailed analysis. A 5-year average of the American Indian or Alaska Native data set will satisfy almost all the requirements. A 3-year average could be used for Chinese and Filipinos, but 5 years are probably necessary for the other API subgroups.

NHANES has a very large sample of Mexican-American and averaging over time will permit fairly detailed cross-classification analyses. The sample was deliberately set up with multi-year averages in mind. None of the other minority subgroups would be helped enough for even simple prevalence rates to have adequate precision.

Table 4-1. Effective sample sizes for Hispanic subgroups, using combined years of data

Data set	Total	Mexican-American	Puerto Rican	Cuban	Central or South American	Other Hispanic
<b>CPS—March</b>						
2 years	12,537	7,727	1,324	523	1,875	1,086
3 years	16,891	10,411	1,784	704	2,527	1,463
5 years	24,773	15,269	2,617	1,033	3,706	2,145
<b>CPS—Monthly</b>						
2 years	6,274	3,863	663	262	940	546
3 years	8,453	5,204	893	353	1,267	736
5 years	12,398	7,633	1,310	518	1,858	1,079
<b>NHIS</b>						
2 years	24,654	15,441	2,620	1,298	3,495	1,802
3 years	33,217	20,804	3,530	1,748	4,709	2,428
5 years	48,718	30,512	5,178	2,564	6,907	3,561
<b>NIS</b>						
2 years	6,232	4,534	511	127	676	386
3 years	8,397	6,109	689	171	911	520
5 years	12,316	8,960	1,010	251	1,337	762
<b>NHANES</b>						
2 years	3,164	3,000	48	20	64	32
3 years	4,746	4,500	72	30	96	48
5 years	7,910	7,500	120	50	160	80
<b>MEPS</b>						
2 years	7,480	5,080	835	314	1,064	187
3 years	10,078	6,844	1,125	423	1,433	252
5 years	14,781	10,039	1,650	620	2,103	370
<b>MCBS</b>						
2 years	705	386	63	102	78	75
3 years	950	520	86	137	106	101
5 years	1,393	762	125	201	155	149
<b>NHSDA</b>						
2 years	3,796	2,406	401	160	548	282
3 years	5,114	3,242	540	216	738	380
5 years	7,501	4,755	792	317	1,082	558

NOTE: The sample cases for each data set reflect the population coverage of the respective surveys. For example, CPS—March covers all persons in the civilian noninstitutional population, whereas NSFG covers women 15 to 44 years of age. The descriptions of the respective data sets note the appropriate population coverage.

Table 4-2. Effective sample sizes for API subgroups, using combined years of data

Data set	Total API	Chinese	Filipino	Japanese	Asian Indian	Korean	Vietnamese	Hawaiian	Other
<b>CPS-March</b>									
2 years	5,072	1,107	947	573	551	539	418	139	630
3 years	6,833	1,492	1,276	792	743	727	563	187	848
5 years	10,022	2,188	1,871	1,132	1,089	1,066	825	274	1,244
<b>CPS-Monthly</b>									
2 years	5,072	1,107	947	573	551	539	418	139	630
3 years	6,833	1,492	1,276	772	743	727	563	187	848
5 years	10,022	2,188	1,871	1,132	1,089	1,066	825	274	1,244
<b>NHIS</b>									
2 years	4,063	934	800	441	396	423	441	139	489
3 years	5,474	1,258	1,078	594	533	569	594	187	659
5 years	8,029	1,848	1,581	871	782	835	871	274	967
<b>NIS</b>									
2 years	1,506	341	292	175	169	165	129	43	192
3 years	2,030	459	394	236	227	222	173	59	259
5 years	2,977	673	578	347	333	327	254	86	380
<b>NHANES</b>									
2 years	226	54	43	26	25	25	19	6	28
3 years	340	81	65	39	38	38	29	10	42
5 years	566	134	108	64	63	63	48	16	69
<b>MEPS</b>									
2 years	596	120	135	50	89	77	35	13	79
3 years	804	162	182	67	107	92	43	18	107
5 years	1,178	237	267	99	176	152	69	26	156
<b>MCBS</b>									
2 years	229	52	43	27	25	25	20	7	28
3 years	308	70	59	36	34	34	27	9	38
5 years	452	102	86	53	50	50	40	13	56
<b>NHSDA</b>									
2 years	531	120	102	62	58	58	45	15	68
3 years	716	162	137	83	79	79	61	20	92
5 years	1,049	238	201	122	116	116	89	30	135

NOTE: The sample cases for each data set reflect the population coverage of the respective surveys. For example, CPS-March covers all persons in the civilian noninstitutional population, whereas NSFG covers women 15 to 44 years of age. The descriptions of the respective data sets note the appropriate population coverage.

Table 4-3. Effective sample sizes for American Indians or Alaska Natives using combined years of data

Data set	American Indian or Alaska Native
CPS—March	
2 years	1,782
3 years	2,401
5 years	3,521
CPS—Monthly	
2 years	1,782
3 years	2,401
5 years	3,521
NHIS	
2 years	1,089
3 years	1,467
5 years	2,152
NIS	
2 years	591
3 years	797
5 years	1,168
NHANES	
2 years	47
3 years	71
5 years	118
MEPS	
2 years	299
3 years	403
5 years	591
MCBS	
2 years	38
3 years	52
5 years	76
NHSDA	
2 years	125
3 years	169
5 years	248

NOTE: The sample cases for each data set reflect the population coverage of the respective surveys. For example, CPS covers persons in the civilian noninstitutional population, whereas NSFG covers women 15 to 44 years of age. The descriptions of the respective data sets note the appropriate population coverage.

The Mexican-American samples in the NIS, MEPS and NHSDA, are fairly large and even 2-year combinations will permit fairly detailed cross-classifications. Five-year combinations are necessary for most of the other Hispanic subgroups. Five years will permit simple analyses of NIS in most of the API subgroups and for American Indians or Alaska Natives. However, even 5 years is not sufficient for the API subgroups and American Indians or Alaska Natives for MEPS and NHSDA. The MCBS sample of minorities is so small that 5 years fails to satisfy most of the precision requirements, except for Mexican-Americans, for whom simple distributions are possible, but not detailed cross-classifications.

#### **4.5 CPS Labor Force Estimates**

The sample sizes shown for CPS in Tables 4-1 through 4-3, both March and monthly, apply to data obtained in a single month of the year. They include the March supplements--income, mobility, work experience, and several other items--and the supplemental information covered in other months, particularly school enrollment and fertility, and voting and registration, which is included every other year. However, CPS collects labor force status each month with the sample size shown for CPS Monthly.

Estimates of annual averages of such items as employment, unemployment, occupation, industry, and related labor force items can be produced by combining data for the 12 months of each year. There is a precedent for such annual averages; for many years CPS has produced annual unemployment rates for the larger states.

The number of observations for annual averages are 12 times the numbers for CPS monthly shown in Tables 3-3 to 3-5, but the effective sample size is lower. The CPS rotation pattern retains households in the sample for a sequence of 4 months, drops them for the next 8 months, and then reinstates them for another 4-month period. As a result, over the course of a year there are multiple observations on most of the sample persons. Furthermore, in the months when a group of sample persons is dropped, most of the sample replacements are neighboring households whose characteristics are usually correlated with the households they replace.

The correlations vary greatly among the labor force items. They are very high for items that tend to persist for most persons over the course of a year, e.g., whether or not in the labor force or employed and for occupation. They are more moderate for unemployment. The U.S. Census Bureau has estimated both the correlations and the effective sample sizes for CPS annual averages.<sup>1</sup> The results indicate that the effective sample size for annual estimates of the unemployment rate is five times the monthly sample. For most of the other labor force items, the effective sample size is only twice the monthly sample. Estimates of average annual unemployment rates, thus, will be based on effective sample sizes five times as large as the numbers in Tables 3-6 to 3-8. Estimates of unemployment rates will satisfy reasonable precision requirements for almost all the minority subgroups. The cost of obtaining annual averages will be quite low since public use files are available.

---

<sup>1</sup> "Current Population Survey Variance Properties" by Gunlicks, Corteville, and Mansur. Proceedings of the Survey Research Methods Section of the 1997 American Statistical Association annual meetings.

## 5. COMBINING RESULTS FROM SEVERAL SURVEYS

As mentioned earlier, a few items are included in several different data sets, e.g., health insurance is covered in NHIS and SIPP and key basic demographic variables such as age, sex, and household relationship are included in almost all surveys, as are certain social and economic items, including income and educational attainment. It is possible to improve estimates of these items for minority subgroups by combining the results from the various surveys requesting similar information. The sample sizes for most subgroups are already fairly high in NHIS and SIPP, and estimates of health insurance from the combined data sets would reduce the sampling errors even more and thus permit the analysis of subgroups, such as specified ages or geographic divisions. It should be noted that health insurance is measured somewhat differently in the surveys, and it is not clear whether the increase in sample size from the combinations of surveys compensates for the problems arising from differences in question wordings that exist.

Even greater reductions in sampling errors are possible for the basic demographic and related characteristics that appear in almost all surveys since they are considered essential covariate items. However, we doubt that it is necessary. These items will be covered in the ACS, which the U.S. Census Bureau expects to initiate in the next few years. The ACS sample will dwarf the samples of the other government surveys, so that it seems sensible to base the analysis of such items as age distributions, income, education, geography, etc. on the ACS. Including the other surveys would hardly reduce the sampling errors. Secondly, the ACS data would not be subject to procedural differences among surveys, e.g., slightly different question wordings, variation in response rates, etc., as would be the case with a combined data set. The ability to improve statistics by combining data from a number of surveys is essentially restricted to a handful of items. Clearly, most information collected in NHIS is not repeated in SIPP or in the other surveys, and the same situation exists in other pairs of data sets. Analysis of the broad array of data items in a survey cannot be improved by combining surveys, unlike the improvements possible by averaging over time.

The MEPS and NSFG samples are subsets of persons in NHIS and there are advantages to combining NHIS data with information from the two surveys, e.g., crossclassifying MEPS or NHIS data with selected NHIS variables, or using NHIS as a source of controls for poststratification. We do not discuss these uses of combinations of surveys in this report because they do not dramatically contribute to the ability of the surveys to provide reasonably reliable data for subpopulations.

## 6. SURVEY SUPPLEMENTATION

### 6.1 Surveys Requiring Sample Supplementation

Sample supplementation involves planning the sample design, carrying out household screenings, and data collection. It will require substantial efforts and costs and obviously should be used only as a last resource, that is when combining years is inadequate. We will use the results from Section 4 to identify surveys needing supplementation. However, we emphasize that such decisions cannot be fully made without careful consideration of the precision needed for specific policy purposes, which is beyond the scope of this report. Consequently, the surveys below and the subgroups for which supplementation is required should be thought of as suggestive rather than as actual recommendations.

We will mostly use the information in Section 4 of this report to identify surveys requiring supplementation. For periodic, rather than annual surveys, Section 3 describes the ability of each survey to provide useful data on minority subgroups.

Supplementation obviously is not required for the National Vital Statistics data sets. Census 2000, and the ACS. Averaging over time in CPS and NHIS, up to 5 years, would provide reasonable statistics for most of the minority subgroups. However, the samples for a few of the smaller subgroups (e.g., Cubans and Hawaiians) would still be quite small, and if those populations are of particular interest, then supplementation is needed. The NHIS Hispanic sample is even larger than CPS so that with the exception of Cuban-Americans, multi-year averages would provide data for adequate precision. The NHIS samples of American Indians or Alaska Natives and APIs are about the same as CPS and multi-year averages will be sufficient for most analytic purposes.

Five-year averages in NIS, NHSDA, and MEPS are sufficient for all Hispanic subgroups, with the possible exception of Cubans, but even 5 years will satisfy only minimal standards for the API subgroups. The American Indian or Alaska Natives sample will permit moderately detailed analysis in NIS but not in NHSDA or MEPS.

The surveys requiring supplementation for all, or almost all, minority subgroups are: NSFG, NHANES, MCBS, NHES, ECLS-B, and ECLS-K.

## 6.2      Designs for Sample Supplementation

Sample supplementation for small subgroups of the population is generally very expensive. This is not due just to the additional interview and data processing costs, but more so to the effort and cost involved in identifying a probability sample of each subgroup of interest. For example, Cuban-Americans constitute one-half of one percent of the U.S. population, so that with purely simple random selection about 200 households have to be screened to locate a single Cuban household. Most of the API subgroups are even smaller and will require even greater screening. This implies screening of hundreds of thousands of households to locate samples of 1,000 or so supplemental cases. Such an effort could cost several millions of dollars for each survey, depending on the amount of supplementation and the desired level of reliability for the smallest subgroups such as Hawaiian and Vietnamese.

Under some circumstances, it is possible to avoid, or reduce, the very great screening effort. The conditions that permit such reductions are described below.

The samples for two surveys are drawn from sampling frames that show race/ethnicity for each person on the frame. ECLS-B is selected from birth records. The vital statistics records contain the detailed race/ethnicity for almost all births. A few of the smaller API subgroups are only identified as "other API" in states that contain only a small percentage of these subgroups, but Chinese, Japanese, Hawaiian, and Filipinos are reported everywhere, as are all of the Hispanic subgroups. Thus, there would be relatively little additional cost to identify a supplemental sample for ECLS-B, although interviewing and data processing costs might still be substantial, depending on the size of the sample supplementation.

A little more effort would be required to supplement MCBS, but it could be done reasonable efficiently. The MCBS sampling frame consists of Medicare beneficiaries in HCFA files. Race and ethnicity are recorded on this file, but not in the detail required. There is a single code for Hispanics and one code for API. Sample supplementation would require selecting a sample of Hispanics and API, screening the sample (possibly by telephone when listed numbers are available) and subsampling persons within each subgroup. More work is involved than for the ECLS-B, but it can be carried out without excessive cost. American Indians or Alaska Natives are also identified on the MCBS frames so supplementation of this population would be similar to that for ECLS-B.

The sampling frames for the other surveys are mostly area segments, although CPS and SIPP are based on census address lists and NHES and NIS use random digit dialing. In these surveys, the race/ethnicity of the sample households are not known in advance of the household contact, and a screening operation is necessary to identify the units eligible for the supplemental sample. Research on possible methods of reducing screening for samples of relatively rare population subgroups was carried out as part of the development of NHANES III procedures. No single procedure appeared to be universally applicable, but substantial gains in efficiency in sampling for Hispanics was possible by oversampling areas with heavy concentrations of Hispanics reported in the most recent census.<sup>1</sup> Further research carried out jointly by Westat and NCHS statisticians confirmed these results and indicated the oversampling rates that would provide the lowest sampling errors.<sup>2</sup> Unfortunately, the research indicated that only trivial improvements were possible through geographic oversampling for APIs or American Indians or Alaska Natives, since relatively high proportions of these populations reside in homes that are scattered throughout the general population. The research described above dealt with the broad race/ethnic groups—Hispanics, APIs, and American Indians or Alaska Natives—and did not explore the detailed subgroups. It is likely that geographic oversampling will be almost as effective for most Hispanic subgroups as for total Hispanics. It is possible that a few of the API subgroups are sufficiently clustered for this kind of a sample to be effective, but a more detailed examination would be necessary to determine this fact. In any case, important gains are not possible for most of the API subgroups, or for American Indians or Alaska Natives. For the Hispanic subgroups, even with the gains in efficiency, a sizeable amount of screening would still be necessary.

Members of subpopulations identified through the NIS screener could be asked question modules addressing topics of interest to ASPE. This is the plan formulated by NCHS for the proposed state and local area integrated telephone survey (SLAITS). The NIS annual screening sample is so large that sufficient sample sizes of each subpopulation can be identified yearly; screening costs would be minimal for such data collection efforts. The respondents, of course, would be limited to households with telephones.

The sample design and estimation method used in the Hispanic Health and Nutrition Survey (HHANES) is a useful precedent to consider for sample supplementation. HHANES did not attempt to sample the entire target population that consisted of Mexican-Americans, Cubans,

---

<sup>1</sup> "Evaluation of Design Options in NHANES '97," report prepared by Westat, May 31, 1994.

<sup>2</sup> "Geographic Oversampling in Demographic Surveys of the U.S.," Waksberg, Judkins and Massey, Survey Methodology, June 1997.

and Puerto Ricans. The HHANES sample was restricted to geographic areas (counties and blocks) containing high concentrations of these subgroups. The sampling frame used for sample selection of PSUs in the Mexican-American sample was restricted to counties with moderate or large numbers of Mexican-Americans or where they constituted reasonably large percentages of the total population. Similarly, the within-PSU sample excluded census block groups or enumeration districts with small numbers of Mexican-Americans. Similar exclusions applied to the Cuban and Puerto Rican samples. The areas in the sampling frames contained well over 80 percent of each subgroup. A model was used to extrapolate the results of the surveys to the total region the data were intended to represent (Southwest for Mexican-Americans, Dade County for Cuban-Americans, and New York City and selected surrounding counties for Puerto Ricans.) The model assumed similar health characteristics for persons inside and outside the areas of heavy concentration of minorities, within specific economic and demographic classes.<sup>3</sup>

The HHANES estimates appeared plausible, and users did not report any problems with the data. Of course, the modeling accounted for less than 20 percent of the total so that it was unlikely that even important problems with the model would introduce serious errors in the results. Use of models would be much more uncertain for API subgroups or American Indians or Alaska Natives. In 1990, 37 percent of APIs and 47 percent of American Indians or Alaska Natives lived in areas that were under 10 percent minority. Some years after a census, these percentages will be even greater. A procedure similar to HHANES that avoided excessive screening would probably be restricted to no more than 50 percent of APIs and about 40 percent of American Indians or Alaska Natives. The validity of data from models that account for the remaining 50 or 60 percent of the total is open to question.

The sampling research for NHANES III mentioned earlier also explored the use of other kinds of sampling frames, in particular, telephone listings of households with Spanish surname, or distinctive names for other minority groups, and subscribers to foreign language newspapers or magazines. None had high enough coverage to be useful.

---

<sup>3</sup> "Estimation in the Southwest Component of the Hispanic Health and Nutrition Examination Survey." 1982-84. by Gonzalez, Ezzati, Lago and Waksberg. Proceedings of the SRMS of the American Statistical Association. 1985.

### 6.3 ACS as a Sampling Frame

At this point, it seems appropriate to note the importance of the ACS as a potential source of information for any or all of the individual population groups of interest. Each is identified, recorded, and entered on the ACS data file. If, as expected, the ACS becomes operational in 2003, it will completely—and virtually immediately—obviate any need for “other” sources of information for those characteristics regularly included in the ACS as core items. To the extent that the ACS also includes periodic, supplementary modules covering the full ACS sample, those data will provide ample sample sizes for each of the population subgroups of interest. Finally, if the need for specialized data exists and cannot be met by any of the approaches described in this report, the ACS lends itself as an efficient and timely source of sample for the subpopulations to be included in a new inquiry, either through supplementary questions added to the core, or the inclusion of a full module in one or more months of interviewing. If those approaches prove infeasible, a separate inquiry can be initiated, using a recent sample previously included in the ACS.

The U.S. Census Bureau has strict confidentiality rules, and it will be difficult, if not impossible, for other statistical agencies to gain access to names and addresses in the ACS. The U.S. Census Bureau has stated that, under its authorizing legislation (Title 13, U.S.C.), it cannot legally make available any personal information collected under Census authority, including names or addresses. In effect, this means that only the U.S. Census Bureau can conduct the interviews (or carry out the measurements) for which the ACS provides a sampling frame. Thus, in attempting to supplement samples for existing inquiries, the statistical agencies responsible for the various surveys will have to decide whether joint responsibility—one contractor conducting most of a survey and the U.S. Census Bureau carrying the same functions for the sample supplement—is operationally feasible.

## 7. SUMMARY OF FINDINGS

- Most of the databases show the race/ethnic identification of each person in sufficient detail to permit subgroup analysis, but the full detail is missing in a few surveys. All statistical agencies are expected to convert to the new race/ethnic classifications within the next few years. Thus, this would be the appropriate period in which to attempt to get uniformity in the detailed race/ethnicity codes to be entered in the data records, if ASPE believes this would permit useful improvements in Federal statistics.
- The National Vital Statistics data sets and the 100 percent data from the decennial censuses are not subject to sampling errors for descriptive analyses, and there are therefore no impediments to subgroup analysis. The long form data in Census 2000 and the ACS are based on such large samples that analyses could be carried out on even very small subgroups with the results subject to only trivial sampling errors.
- None of the other surveys provide sufficient precision to permit sophisticated analysis of all subgroups. The larger data sets —CPS-March, NHIS, and NHES—contain adequate samples of Mexican-Americans, and analyses based on cross-classification are possible. However, only simple distributions could be carried out reliably for most of the other race/ethnic subgroups. CPS-Monthly, SIPP, NIS, and MEPS also provide satisfactory data for Mexican-Americans, but even simple distributions for most of the other subgroups would have poor reliability. In the other surveys, only limited analysis of some of the larger subgroups could be carried out with any confidence.
- Multi-year averages, of course, would improve the precision. Five-year averages will provide samples large enough to satisfy analytic needs for most Hispanic subgroups for the larger data sets, i.e., CPS and NHIS. Three-year averages in the current NHANES would provide reasonably precise data for Mexican-Americans, and 5 or 6-year averages would permit analyses of detailed age-sex classes. However, in all surveys the Cuban sample and the samples of most of the API subgroups would still be too small for anything but simple analyses. The other data sets would also be improved by averaging over time, but the effective sample sizes of many subgroups would still be small.
- It is probably not practical to obtain multi-year averages for the periodic (as distinct from annual) surveys. These comprise NSFG, SIPP, ECLS-B, and ECLS-K. We also include NHES in this category since, although it is annual, the main data content varies from year to year.
- Annual averages of unemployment rates for each subgroup in CPS would have reasonable precision and could be obtained with relatively little effort. Annual averages for other labor force items would be only a little better than monthly statistics.
- There are a few items that appear on more than one survey, and combining the results would improve precision. However, this is a fairly rare occurrence and can satisfy only limited data needs.
- If the U.S. Census Bureau goes ahead with its plans for the ACS (currently scheduled to start in the year 2003), it could be a major resource for subgroup analysis. First, the ACS will be able to supply annual statistics on a variety of demographic, social, and economic

characteristics for each subgroup. Secondly, it could become the vehicle for obtaining much needed information for these groups, either through the addition of questions to the ACS, or through a special effort which used the ACS as a source of sample. Finally, it could become the sampling frame for the selection of supplemental samples for other surveys, substantially reducing the cost of sample supplementation. However, in such cases, a number of bureaucratic hurdles would have to be overcome. Whether this could be done to the satisfaction of both the U.S. Census Bureau and the sponsoring agencies is uncertain.

- Sample supplementation for most surveys will be quite expensive if use of the ACS is not practical. Statisticians have developed devices for reducing the sampling and screening costs for small population groups, but a considerable amount of screening would still be required. Also, it is unlikely that the devices would be effective for all subgroups.
- We would like to repeat the caveats mentioned earlier in this report:
  1. The sample sizes provided in Tables 3-3 to 3-5 which were used to estimate effective sample sizes and to ascertain whether surveys achieved reasonable standards of precision, refer to specific time periods (reported in Section 1.2). The samples in most Federal surveys are fairly stable, but changes are made from time to time. Although small changes in sample size in the order of 10 or 15 percent will have only a negligible effect on the conclusions drawn in this report, much larger revisions occasionally occur. Before going ahead with a study of a subgroup in a particular survey, the analyst should refer to the documentation for the survey to see whether the sample sizes in Tables 3-3 to 3-5 are still applicable. Any important changes in the sample should be taken into account.
  2. The sample sizes in this report refer to each survey's total sample for the race/ethnic subgroup. When the analysis is restricted to a subclass of the total (e.g., all males, or females, persons in a specific age group, etc.) the sample size should be adjusted accordingly.
  3. In a few surveys, a subsample is used for some variables. For example, NHIS frequently collects selected information from only one person in each sample household. Similarly, NHANES uses random subsamples of the full sample for some items. An analyst should ascertain whether or not the full sample is used for the variables of interest, and determine whether the sample sizes in Tables 3-3 to 3-5 are appropriate.
  4. The design effects reported in Table 3-2, which are necessary for the estimation of effective sample sizes, are averages over a broad set of items, and reflect variables for which correlations among household members are not excessive. There are some items for which almost all household members have the same value, e.g., presence of health insurance, poverty status, urban-rural residence, region of residence. The design effects are much larger for such items. Section 3.5 discusses methods of dealing with such situations.

5. Finally, it is important to recognize that considerable "noise" is to be found in the statistics. For example, small differences in reporting of race/ethnicity among some of the databases, minor variations in sample size from year to year even when there are no changes in sample design, and the use of average design effects, which do not reflect the variation among items, are all sources of "noise." As a result, the conclusions drawn in this report should be considered as approximations, but are sufficiently accurate as to be a useful guide on the kinds of analyses of race/ethnic subgroups that are possible with the various databases.

## **APPENDIX A**

### **SOURCES USED TO PROVIDE ESTIMATES OF DESIGN EFFECTS**

Table A-1. Sources used to provide estimates of design effects

Survey	Source
Census	
Census 2000	Estimated from description of sample design
ACS	Estimated from description of sample design
CPS	U.S. Census Bureau reports, Series P-60, No's 198 and 200
SIPP	Communication from U.S. Census Bureau
NCHS/CDC	
NHIS	NHIS report on Variance Estimation, by D. Judkins and D. Wright, September 20, 1990
NSFG	National Survey Family Growth, Cycle IV, Evaluation of Linked Design, NCHS Series 2, No. 117, July 1993
NIS	Estimated from description of sample design
NHANES	Sample Design Research for NHANES '97, Oct. '94, page 2-1, combined with description of sample design
AHRQ	
MEPS	Design Effects of Survey Estimates Derived from the 1996 MEPS, William Yu, paper prepared for 1999 annual meetings of the American Statistical Association
HCFA	
MCBS	Components of Variance and Nonresponse Adjustments for MCBS, by D. Judkins and A. Lo. Proceedings of Section on Survey Research Methods of the A.S.A., 1993
SAMHSA	
HSDA	Methodological Resource Book for the 1991 NHDAS
NCES	
NHES	Unit and Item Respondent Weighting and Imputation Procedures in the 1995 NHES, NCES Working Paper 97-06. Estimated from description of sample design.
ECLS-B	Estimated from description of sample-design.
ECLS-K	Estimated from description of sample-design.

